AD_____

Award Number: W81XWH-04-1-0512

TITLE: Promoter and Cofactor Requirements for SERM-ER Activity

PRINCIPAL INVESTIGATOR: Jason Carroll, Ph.D.

CONTRACTING ORGANIZATION: Dana Farber Cancer Institute
Boston, Massachusetts 02115

REPORT DATE: May 2006

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2006	Annual (15 Apr 05 - 14 Apr 06)

4. TITLE AND SUBTITLE
Promoter and Cofactor Requirements for SERM-ER Activity

5. FUNDING NUMBERS
W81XWH-04-1-0512

6. AUTHOR(S)
Jason Carroll, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dana Farber Cancer Institute
Boston, Massachusetts 02115

E-Mail: jason_carroll@dfci.harvard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

The mechanism by which Estrogen Receptor regulates transcription is poorly understood. We aimed to identify, in an unbiased manner, DNA elements and proteins that augment or inhibit ER activity in vivo. We used a combination of Chromatin Immunoprecipitation (Chip) with various novel approaches to identify contributing factors. We initially mapped ER binding using Chip in combination with tiling microarrays that cover the entire non-repetitive sequence of chromosomes 21 and 22. This study resulted in several novel conclusions, including the fact that ER rarely binds to promoter regions but uses distal enhancer and that these distal enhancers are defined by the presence of a Forkhead protein, namely FoxA1. We also showed that FoxA1 is essential for ER binding to the chromatin and for subsequent gene transcription. More recently, we have extended on these findings and have mapped all in vivo ER and RNA Polymerase II binding sites in a breast cancer cell line. We are currently mining this large dataset, but have already found a number of novel ER associated proteins that appear to be essential for ER-mediated transcription

14. SUBJECT TERMS	15. NUMBER OF PAGES
Estrogen Receptor, transcription, chromatin, SERM	28
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Estrogen and the Estrogen Receptor (ER) regulate gene transcription that ultimately impinges on cell division and cancer progression, but the mechanisms are poorly understood. These limitations in our knowledge are primarily a function of the inadequate molecular and biochemical tools available to analyze ER transcription on a genome-wide scale. Most work to date has focused on *in vitro* systems for assessing requisite promoter regions in reporter assays, although it is becoming apparent that fragments of DNA function completely differently *in vitro*, when compared to chromatin contexts. As such, the proposal of this fellowship was to address *in vivo* mechanisms of protein-chromatin interactions by using the power of Chromatin Immunopreciptations in combination with novel approaches for identifying regulatory components. We initially attempted to identify proteins bounds to promoters of interest in a chromatin context, although this proved to be unsuccessful, primarily due to the conclusion that fragments of DNA (approximately 1 kb regions) integrated into random regions of the chromatin do not function in a transcriptional manner, under any experimental conditions. This finding severely restricted the ability to use this specific cell system approach for further analysis. However, another approach to assess protein-chromatin interacts was taken, in which ChIP was again employed, but was used to identify ER binding sites in an unbiased manner which was subsequently analyzed to find interacting proteins. This novel approach combined ChIP with microarrays covering all of chromosomes 21 and 22 at 35 bp resolution to map ER binding sites on a chromosome-wide level in order to reveal the underlying DNA and protein components of the ER transcriptional machinery. Furthermore, we extend on this to map all ER and RNA PolII binding sites across the entire human genome of a breast cancer cell line.

Body

Previously reported work

The ultimate goal of the project was to identify novel proteins that interact with the ER complex during transcription, using in vivo Chromatin Immunoprecipitation (ChIP) assays with novel approaches for identifying proteins. We initially aimed to generate MCF-7 (breast) and ECC1 (endometrial) cancer cells with a single Lox-Luciferase integration cassette embedded within the chromatin, that could be used as an entry point for introduction of promoters of interest. These promoters, included c-Myc, EBAG9, TFF-1 and IGF-1, would be assessed for transcriptional activity (as assessed by luciferase activity) and this transcriptional activity could be monitored when various mutants of the promoter sequences were re-introduced into the same locus of the chromatin. These promoters had previously be cloned into luciferase reporter assays and shown to possess potent transcriptional activity in this histone-free in vitro assay. The secondary goal was to tag the promoters of interest and to subsequently use the tag to precipitate the DNA and assess what proteins are associated with it, in order to identify, in an unbiased manner, the proteins that bind with ER and potentially function as coactivators to augment transcription. We previously reported that we had generated several MCF-7 clonal cell lines and ECC1 clonal cell lines and screened them for the presence of a single integration site. Furthermore, we generated the cloning vectors required for introduction of various promoter regions of interest into the chromatin. We performed these experiments and selected clonal cell lines that contained c-Myc, EBAG9, TFF-1 and IGF-1 promoter regions, to establish individual cell lines that had the different promoter regions in the same chromatin context. However, when we assessed luciferase activity in any of the cell lines, we could not detect any transcription activity under any conditions, including hormone depletion, estrogen addition and growth factor stimulation. This was the case for all the different clonal cell lines and suggested that either the cassette had integrated (in all cases) into a region of the chromatin that was not conducive to transcriptional activity, or alternatively that the 1kb promoter regions could not induce transcription in these chromatin conditions. To identify the mechanisms for this failure of transcriptional activity, we introduced the CMV promoter sequence into the Lox-

integration site in the chromatin and select cells to generate stable clonal cell line that contained the potent CMV promoter in the Lox-Luciferase cassette. When we assayed for luciferase activity using this powerful promoter, we could not detect activity in any MCF-7 clones and only marginally detected activity in one ECC1 clonal cell line. This suggested that in a chromatin context, small DNA sequences with in vitro activity cannot function appropriately. In order to establish if new clonal cell lines could be derived that contained the random Lox-luciferase cassette integrated into a more euchromatic regions that may be more permissive of transcription, we re-transfected in the Lox-Luciferase cassette, selected cells, generated clonal cell lines and assessed them for activity by recombining the CMV promoter into the Lox-luciferase site. None of the newly generated clones possessed any transcriptional activity, negating the ability of this approach to assess the transcriptional activity from specific piece of DNA. Due to this limitation, it was no longer possible to pursue the later aims of identifying essential DNA motifs for transcription and subsequently identifying novel cofactors during ER-mediated transcription. To circumvent this problem we attempted to achieve the same original goal by combining ChIP with microarrays that cover significant regions of unexplored sequence in order to find genuine in vivo ER binding sites that could subsequently be mined to find enriched DNA binding elements and shed light on the unknown cofactors that augment ER transcription.

Development and validation of ER ChIP and amplification of DNA

MCF-7 breast cancer cells are used as a model to understand ER action. We grew MCF-7 cells in complete media and subsequently depleted them of serum by treating for 3 days in Charcoal Dextran Treated (CDT) media. This hormone depleted media results in cell cycle arrest, which was assessed by flow cytometry. Estrogen was added for increasing time periods and the cells were fixed in formaldehyde to maintain protein-protein and protein-DNA interactions, after which chromatin was collected and a specific antibody to ER was used to immunoprecipitate ER, the associated proteins and interacting DNA fragments. The DNA was purified and real time PCR was performed using primers against the promoter of TFF-1, a well-characterized estrogen target gene. The data was normalized to DNA content and further normalized to total genomic DNA (Input) to

assess the enrichment of TFF-1 promoter bound to ER at the different time points of estrogen treatment. A cyclic association of ER was observed, with a maximal recruitment of ER at 45 minutes.

We used DNA bound to ER at the 45 minute time point as a source of chromatin to identify ER binding sites. Due to the low yield of DNA during ChIP (approximately 1 to 2ng), but the large amount of DNA required for microarray analysis (several ug), DNA amplification was required. We utilized a ligation-mediated PCR approach (LM-PCR) that involved a number of steps: 1. Validated DNA was end filled to generate blunt ends, 2. pre-annealed linkers were ligated onto the ends of the DNA fragments in a random manner to generate similar ends on each DNA fragment, 3. limited PCR was performed using a primer against the linker region to amplify the DNA, 4. DNA was purified, quantitated and validation of enrichment was performed using TFF-1 as a positive control. Once the DNA was assessed and shown to be abundant with maintenance of ER binding enrichment on tested sites, we end labeled the DNA using dNTP-biotin and prepared the samples for microarray hybridization.

ChIP-on-chip discovery of ER binding sites and interacting proteins on chromosomes 21 and 22

The microarrays used were generated by Affymetrix and cover the entire non-repetitive DNA sequences of chromosomes 21 and 22 using 25 bp probes every 35 bp across the entire chromosomes. This results in approximately 1 million probes that cover 35 million bp, including all the genes, introns, and intergenic sequences of chromosomes 21 and 22. These probes are split on a 3 microarray set in order to cover this large region of the genome. As a positive control, TFF-1, the previously validated estrogen target gene is located on chromosome 21. The DNA associated with ER by ChIP was hybridized to the microarrays and data was analyzed by comparing the signal from each Perfect Match (PM) probe and control Mismatch (MM) probes. Once this difference was established, non-parametric Wilcoxin ranked sum analysis was performed using a sliding window of 600bp to identify clusters of positive probes that represent ER binding sites. This analysis involves some simple parameters, which included the requirement for multiple adjacent

probes to be positive and for gaps of a maximum size to limit peak identification. This resulted in 57 ER binding sites on chromosomes 21 and 22. As an example, we found ER binding at the promoter of TFF-1, exactly 400bp upstream of the transcription start sites, where a well defined ERE was located (Figure 1). Surprisingly however, we also found an ER binding site 10.5 kb upstream from TFF-1 gene suggesting it may be an enhancer.

To validate some of the newly identified ER binding sites, we designed primers against the chromosomal co-ordinates that were defined as ER binding site peaks and performed standard ER Chip followed by real time PCR of the newly identified sites. All of the sites we tested proved to be genuine in vivo ER binding sites, confirming the power of the ChIP-on-chip approach. We found unique ER binding patterns near several genes of interest, including 10 ER binding sites in the middle of the DSCAM-1 gene, 6 ER binding sites more than 150kb from the transcription start site of the Nuclear Receptor cofactor, NRIP-1, and 3 ER binding sites 15-25 kb upstream of the XBP-1 transcription factor. All of these genes were shown to be estrogen regulated. Furthermore, we performed ChIP using antibodies against RNA Polymerase II and the ER cofactor AIB-1, both of which were shown to be recruited to the ER binding sites in an estrogen dependent manner. To prove that the ER binding sites that were, in some cases, significant distances from the putative gene targets, we applied a Chromosome Conformation Capture (CCC) approach to identify long distance cis-regulatory elements, which proved successful in two of the three assessed cases, including TFF-1 and NRIP-1. This for the first time confirmed that long distance enhancers are used as primary ER binding sites for transcription.

Using the pool of 57 ER binding sites on chromosome 21 and 22, we screened the sequences for DNA binding motifs that were enriched more than expected by chance and found two such elements, namely an Estrogen Responsive Element (ERE) and a Forkhead motif (Figure 2). The finding of EREs validated the technique and proved that we were in fact finding genuine ER binding sites, but the identification of the Forkhead motif suggested a novel role for Forkhead proteins and ER. A search of all the Forkhead proteins (there are approximately 40 members known, all of which can bind to the same

Forkhead motif that was enriched within the ER binding site) in MCF-7 cells using publically available data revealed the high expression of one Forkhead protein, namely FoxA1, which was also shown to correlate with ER status in breast tumors. Furthermore, FoxA1 was shown by others to bind to other Nuclear Receptors including Androgen Receptor (AR) and Glucocorticoid Receptor (GR), all of which suggested that this was the Forkhead protein most likely to bind Forkhead motifs in our system. We performed ChIP of FoxA1 (as well as several other Forkhead proteins as controls) followed by PCR of a number of the newly identified ER binding sites. This resulted in data showing that FoxA1 binds to approximately 50% of all ER binding sites, but interestingly, unlike most proteins co-operating with ER, FoxA1 was on the chromatin before estrogen addition and dissociates from the DNA after estrogen treatment, coincident with ER loading onto the DNA. Since thousands of predicted ER binding sites (in the form of computationally predicted EREs) occurred on chromosome 21 and 22, but only 57 binding sites were observed, the presence of FoxA1 provided the possibility that this Forkhead protein may dictate where ER can bind to the chromatin. To assess this hypothesis, we designed siRNA against FoxA1 and transfected this siRNA into MCF-7 cells, along with siLuciferase as a control. We subsequently assessed FoxA1 protein levels after siRNA (Figure 3) and collected RNA after vehicle or estrogen stimulation. When we assessed the estrogen induced mRNA changes in several estrogen target genes on chromosomes 21 and 22, we observed a significant decrease in estrogen induction when FoxA1 was silenced, suggesting that the newly identified ER co-operating factor, FoxA1, is essential for ER activity. In order to assess whether FoxA1 was required for ER to bind to the chromatin, we performed siFoxA1 silencing and then assessed ER recruitment to a number of tested sites by ER Chip. We found (Figure 3) that ER could not bind to DNA in the absence of FoxA1, showing a requirement for FoxA1 in defining where and how ER can bind to the chromatin.

ChIP-on-chip discovery of ER binding sites and interacting proteins on the whole human genome

The success of ER Chip-on-chip studies on chromosomes 21 and 22 permitted the identification, for the first time, of a factor that is required to get ER to DNA for

transcription. However, these studies were limited to chromosomes 21 and 22. Technical advances by Affymetrix resulted in the production of the entire human genome tiled at 35 bp resolution of 14 microarrays (6 million probes per chip), which we used in combination with ER and RNA Polymerase II Chip to map all ER and RNA Polymerase II binding sites in the entire genome. This was performed in triplicate and the data was analyzed using a novel bioinformatics approach, entitled MAT, which normalized within each probe, providing exceptional filtering of the data to generate genuine ER binding sites with a false discovery rate of less than 1%. This resulted in 3,665 ER and 3,629 RNA Polymerase II binding sites across the entire genome. Analysis of the ER and RNA Polymerase II sites revealed a significant degree of sequence conservation with the binding sites, suggesting that these discrete regions are conserved in multiple species, highlighting their biological significance during evolution.

To address the major goal of this proposal, we again attempted to identify proteins that would co-operate with ER to mediate transcription, although the current approach used a statistical enrichment of transcription factor binding sites within the newly identified ER binding sites. When we performed this analysis of all 3,665 ER binding sites, we find EREs and Forkhead motifs, as previously identified from chromosome 21 and 22 analyses. However, we also find C/EBP, AP-1 and Oct elements enriched with the ER binding sites, suggesting that the factors that bind to these elements likely contribute, to some degree, to ER transcription. As such, we performed Chip of C/EBPα, Oct-1 and c-Jun (which binds AP-1 motifs) followed by real time PCR of a number of newly discovered ER binding sites. We find C/EBPα, Oct-1 and c-Jun binding to a number of ER binding sites. We designed siRNA to each of these newly implicated factors and showed that by specifically silencing each, we would partially abrogate the estrogen induction of a number of target genes. These data provide novel insight into the proteins that co-operate with ER on the chromatin to regulate gene transcription.

Key Research Accomplishments

- Defined the optimal time period of ER recruitment to promoter of a target gene in vivo using Chip.

- Performed ER Chip-on-chip for the first time, on chromosomes 21 and 22 tiling arrays and validated the findings

- Discovered that ER used long distance enhancers for transcription and that promoter binding is rare

- Show by Chromosome Conformation Capture, that distal enhancers physically interact with the promoters of target genes in an estrogen dependent manner

- Identified FoxA1 as a component of the ER pathway, a factor that is required for ER to bind to DNA

- Performed ER Chip-on-chip using the whole human genome tiled at 35 bp

- Identified all ER and RNA Polymerase II binding sites through-out the entire genome and correlate this with gene expression to define novel patterns of transcription

- Found a number of new ER co-operating factors and show their involvement in ER transcription

Reportable outcomes

- Poster presented at Keystone 2004
- Seminar presented at Project Program Grant retreat 2004
- Seminar presented at Project Program Grant retreat 2005
- Poster presented at DOD conference 2005
- Manuscript published in Cell 2005
- Development of new analysis tool for Chip-on-chip data (MAT)
- First map of ER binding on entire genome
- Invited seminar, Novartis Institute for Biomedical Research 2005
- Invited seminar, Biomedicum, University of Helsinki, Finland 2005
- Poster award at Harvard breast cancer symposium 2005
- Invited seminar at Harvard breast cancer symposium 2005
- Poster presented at Keystone 2006
- Poster presented at Harvard breast cancer symposium 2006
- Manuscript under review at Cell 2006
- Faculty position gained at Cancer Research UK and University of Cambridge, UK. To start 2007

Conclusions

The sum of these data provides significant advances in our understanding of ER transcription. Previous models of ER biology involved ER binding to promoters of target genes, followed by association of cofactors and gene induction. We now show that promoter regions are very rarely ER binding sites, and instead ER docking sites exist significant distances from the target genes. Identification of enriched motifs within ER binding sites provided clues about the factors that may be defining these discrete regions as genuine in vivo ER binding sites out of the thousands of putative binding sites. This analysis led to the identification of Forkhead motifs, which we subsequently showed to be function, in that FoxA1 could associate with heterochromatin and define the ER binding sites. Further analysis of ER binding sites, on a genome-wide level, led to the characterization of the complete ER map, which will be mined for significant time in order to define ER mechanisms of transcription of any target gene of interest. Already, this genome-wide ER binding information has permitted the identification of several new interacting factors, including C/EBP, Oct and AP-1 proteins. Future work will further define this complex network of transcription factors in the estrogen response pathway and will focus on delineating mechanisms by which estrogen can up-regulate some genes and down-regulate others.

References

Carroll, J. S., Liu, X. S., S., B. A., Meyer, C. A., Li, W., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., *et al.* (2005). Chromosome-wide Mapping of Estrogen Receptor Binding Reveals Long-range Regulation Requiring the Forkhead Protein FoxA1. Cell *122*, 33-43.

Carroll, J. S., Meyer, C. A., Song, J., Li, W., Brodsky, A. S., Geistlinger, T. R., Eeckhoute, J., Keeton, E. K., Fertuck, K. C., Hall, G. F., *et al.* (2006). Defining the Estrogen Receptor Cistrome in Breast Cancer. Cell *Under Review*.

Johnson, W. E., Li, W., Meyer, C. A., Gottardo, R., Carroll, J. S., Brown, M., and Liu, X. S. MAT: Model-based analysis of tiling-arrays for ChIP-chip. PNAS *Under Review*.

Appendices

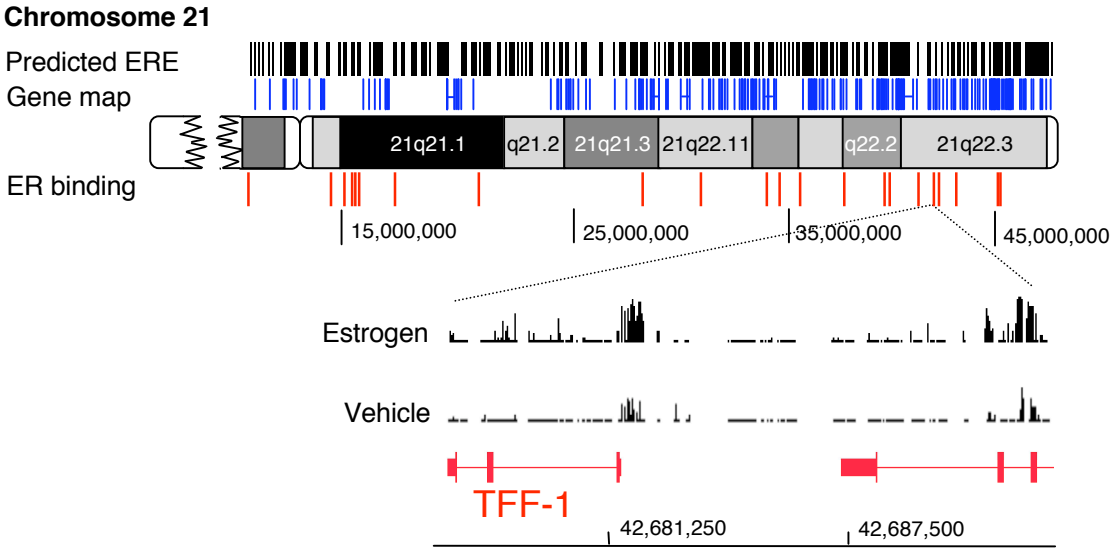

Figure. 1
Map of ER binding sites on chromosomes 21 after estrogen stimulation. Genes locations are shown in blue bars. Gene locations are based on the April 2003 genome freeze in the UCSC browser using Genbank RefSeq positions. Predicted EREs are shown as black bars and ER-binding sites are shown as red bars. An expanded view of the TFF-1 gene region is shown as signal difference between ER ChIP and Input DNA for both the estrogen and vehicle treated cells. The TFF-1 gene is shown in its genuine 3'-5' orientation.

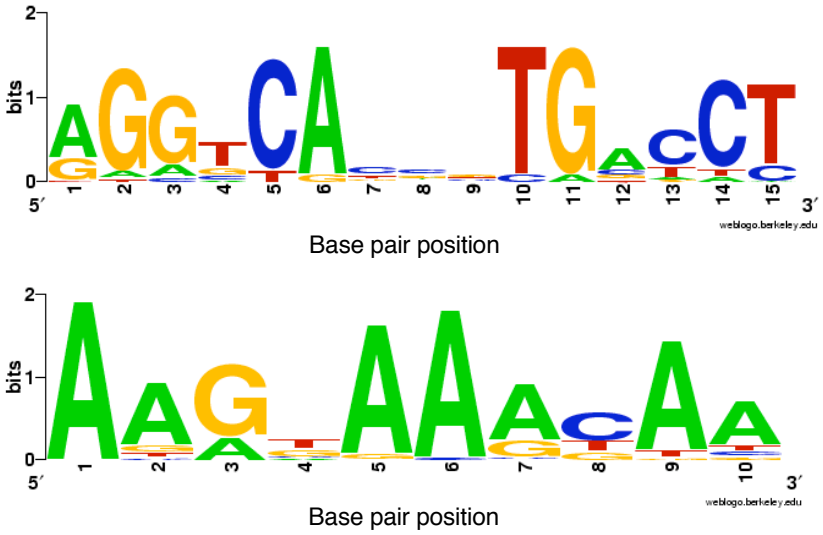

Figure. 2

Presence of enriched motifs within ER binding sites. An unbiased motif screen of all the ER binding sites on chromosomes 21 and 22 revealed the presence of two enriched motifs, an ERE and a Forkhead binding motif, both of which are visually represented in WebLogo (http://weblogo.berkeley.edu).

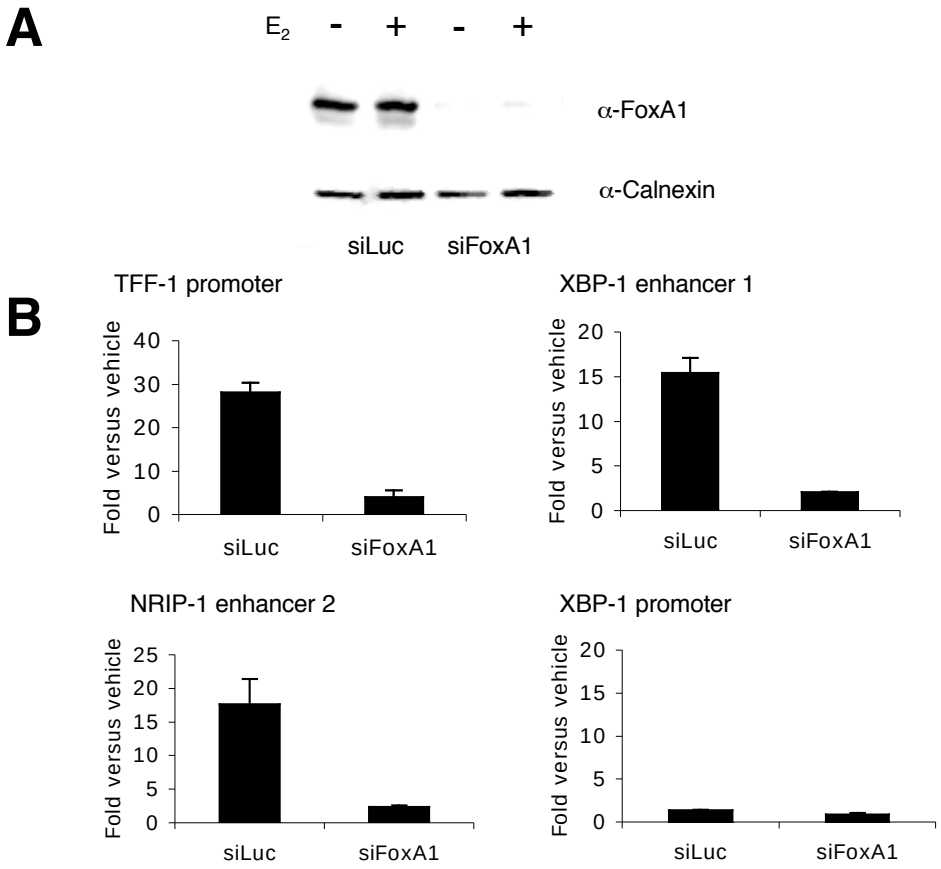

Figure 3 Specific targeted knockdown of FoxA1 and the effects on estrogen-mediated transcription. **(A)** siRNA to FoxA1 was transfected into MCF-7 cells and changes in FoxA1 protein levels were determined. **(B)** ER ChIP was performed after vehicle or estrogen treatment of siLuc or siFoxA1 transfected cells and real time PCR was conducted on TFF-1 promoter, XBP-1 enhancer 1, NRIP-1 enhancer 2, as well as XBP-1 promoter as a negative control. The data are fold enrichment over vehicle treated. **(C)** Changes in mRNA levels of all estrogen-regulated genes on chromosomes 21 and 22. The data are estrogen-mediated fold enrichment compared to vehicle (ethanol) control and are the average of three separate replicates ± Std. Dev.

Cell, Vol. 122, 33–43, July 12, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.05.008

Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1

Jason S. Carroll,[1] X. Shirley Liu,[2,4] Alexander S. Brodsky,[3,5] Wei Li,[2,4] Clifford A. Meyer,[2,4] Anna J. Szary,[1] Jerome Eeckhoute,[1] Wenlin Shao,[1] Eli V. Hestermann,[6] Timothy R. Geistlinger,[1] Edward A. Fox,[3] Pamela A. Silver,[3,5] and Myles Brown[1,*]

[1]Department of Medical Oncology
[2]Department of Biostatistics and Computational Biology
[3]Department of Cancer Biology
Dana-Farber Cancer Institute
Harvard Medical School
44 Binney Street
Boston, Massachusetts 02115
[4]Harvard School of Public Health
Harvard Medical School
677 Huntington Avenue
Boston, Massachusetts 02115
[5]Department of Systems Biology
Harvard Medical School
200 Longwood Avenue
Boston, Massachusetts 02115
[6]Department of Biology
Furman University
3300 Poinsett Highway
Greenville, South Carolina 29613

Summary

Estrogen plays an essential physiologic role in reproduction and a pathologic one in breast cancer. The completion of the human genome has allowed the identification of the expressed regions of protein-coding genes; however, little is known concerning the organization of their *cis*-regulatory elements. We have mapped the association of the estrogen receptor (ER) with the complete nonrepetitive sequence of human chromosomes 21 and 22 by combining chromatin immunoprecipitation (ChIP) with tiled microarrays. ER binds selectively to a limited number of sites, the majority of which are distant from the transcription start sites of regulated genes. The unbiased sequence interrogation of the genuine chromatin binding sites suggests that direct ER binding requires the presence of Forkhead factor binding in close proximity. Furthermore, knockdown of *FoxA1* expression blocks the association of ER with chromatin and estrogen-induced gene expression demonstrating the necessity of FoxA1 in mediating an estrogen response in breast cancer cells.

Introduction

Estrogen is an essential regulator of female development and reproductive function and has been implicated as a causal factor in breast and endometrial cancers. Estrogen-regulated gene expression is mediated by the action of two members of the nuclear receptor family, ERα and ERβ, with ERα being dominant in both breast epithelial cells and in breast cancer. Significant progress has been made over the past decade in defining the complex interactions between chromatin and an array of factors involved in ER-mediated gene expression (Halachmi et al., 1994; Metivier et al., 2003; Shang and Brown, 2002; Shang et al., 2000), including the cyclic association of ER, p160 coactivators (such as AIB-1), histone acetyl transferases (HAT), and chromatin modifying molecules, such as p300/CBP and p/CAF, with target promoters in an ordered temporal fashion (Metivier et al., 2003; Shang et al., 2000).

In addition, a number of recent strategies including gene expression profiling on microarrays have identified potential ER target genes in human breast cancer cells and only a few *cis*-elements targeted directly by ER have been identified to date. For example, estrogen responsive elements (ERE) have been identified within the 1 kb 5′-proximal region of the estrogen-regulated genes *TFF-1 (pS2)*, *EBAG9*, and *Cathepsin D* (Augereau et al., 1994; Berry et al., 1989; Ikeda et al., 2000), and the proximal promoters of target genes that lack EREs, including *c-Myc* and *IGF-I*, contain AP-1 and Sp-1 sites that appear essential for transcription in in vitro reporter assays (Dubik and Shiu, 1992; Umayahara et al., 1994). Few, if any regulatory elements at significant distances from the mRNA start sites of target genes have been shown to be directly targeted by ER, and computation approaches to identify novel ER binding domains have focused primarily on gene proximal regions (Bajic and Seah, 2003; Bourdeau et al., 2004). However, more progress has been made in studies of *β-globin* gene regulation which has contributed to our understanding of general mechanisms of transcriptional regulation and has shown that locus control regions (LCR) up to 25 kb from the gene are capable of enhancing gene transcription (recently reviewed in Bulger et al. [2002]). In this study, we have undertaken an unbiased approach to identify all regulatory regions that may play a role in ER-mediated transcription by combining chromatin immunoprecipitation (ChIP) analyses of in vivo ER-chromatin complexes with Affymetrix tiled oligonucleotide microarrays that cover the entire nonrepetitive sequences of chromosomes 21 and 22, including, importantly, all the intergenic regions. Most previous ChIP-microarray studies have focused primarily on promoter regions (Odom et al., 2004) or CpG islands, which represent promoter-rich sequences (Weinmann et al., 2002). The tiled arrays used here are composed of 25 bp probes located at 35 nucleotide resolution (Cawley et al., 2004; Kapranov et al., 2002) and permit the opportunity to interrogate previously unexplored regions of chromosomal DNA. The 780 characterized or predicted genes on chromosomes 21 and 22 represent about 2% of the total number of genes (Kapranov et al., 2002) and thus provide a representative model for

*Correspondence: myles_brown@dfci.harvard.edu

the unbiased identification of ER-mediated gene regulation paradigms.

Here we find a discrete number of ER binding sites across chromosomes 21 and 22, almost all of which are in nonpromoter proximal regions. We explored underlying biological patterns within the list of genuine chromatin-interacting domains and identified common motifs highly enriched in these regions. Using this information, we prove that the distal ER binding sites are discrete chromatin regions involved in transcriptional regulation and that a Forkhead protein, at these sites, is required for activity by ER.

Results

ER Occupies a Limited Number of Binding Sites on Chromosomes 21 and 22

Estrogen-dependent MCF-7 breast cancer cells were deprived of hormones and stimulated with estrogen or vehicle for 45 min, a time we have previously shown to have maximal recruitment of ER to the promoters of several known gene targets, including *Cathepsin D* and *TFF-1* (Shang et al., 2000). Following ChIP, ER-associated DNA was amplified using nonbiased conditions, labeled, and hybridized to the tiled microarrays. Relative confidence prediction scores were generated by quantile normalization across each probe followed by an analysis using a two-state Hidden Markov model (Rabiner, 1989). These scores included both probe intensity and width of probe cluster. Triplicate experiments eliminated stochastic false positives, after which peaks that reproducibly appeared at least twice in the three replicates were included. Real-time PCR primers were designed against numerous peaks in the list, and directed ER ChIP was conducted to identify the boundary between the true ER binding peaks (>1.5-fold enrichment over input) and the false positives (data not shown) and generate the final list of 57 estrogen-stimulated ER binding sites within 32 discrete clusters (Figures 1A and 1B and see the Supplemental Raw Data in the Supplemental Data available with this article online).

As one example of the validity of this method, the localization of ER to the proximal promoter 400 bp region of the estrogen-regulated gene, *TFF-1*, was observed. A functional ERE had been previously mapped to the region 393 to 405 bp upstream from the transcription start site of *TFF-1* (Berry et al., 1989). Furthermore, a region 10.5 kb upstream of the *TFF-1* transcription initiation site (Figure 1A) was also found to be bound by ER. Interestingly, an estrogen-inducible DNase I hypersensitive site has been previously mapped 10.5 kb upstream from the *TFF-1* start site (Giamarchi et al., 1999), though the region had not been further characterized. Our data now define these regions as authentic ER binding sites.

Within the small list of 57 ER binding sites, we observed 32 ER binding clusters, some of which were proximal to genes previously implicated as estrogen targets, including the transcription factor *XBP-1, DSCAM-1*, and the nuclear receptor coregulator *NRIP-1* (Cavailles et al., 1995; Pedram et al., 2002; Wang et al., 2004). Binding sites were also observed within 200 kb from genes not previously implicated as estrogen targets, including *SOD-1*, a superoxide dismutase gene involved in scavenging oxygen-free radicals (Beckman et al., 1993; Singh et al., 1998) and implicated in tamoxifen-resistant progression in MCF-7 xenografts (Schiff et al., 2000). None of these genes recruited ER to a proximal 5′ promoter region, but possessed divergent patterns of association. The *XBP-1* gene, recruited ER to three distinct and discrete regions 13.2 kb to 22.9 kb upstream of the transcription start site (Figure 1B). *DSCAM-1* contained a clustering of ten intronic ER binding sites, more than 0.5 Mb from the transcription initiation site. *NRIP-1* contained six ER binding sites in a region of chromosome 21 well known for its scarcity of genes (Katsanis et al., 1998). 5′ RACE was performed on *NRIP-1* to determine the exact location of the transcription start site and the distance between the ER binding sites and the genuine transcriptional start site. Sequencing of the 5′ terminus of the *NRIP-1* transcript after estrogen stimulation revealed the presence of two previously missed exons for *NRIP-1*, 74.96 kb and 97.39 kb from the previously annotated gene start site (data not shown). Therefore, the ER binding domains exist 107 to 144 kb from the genuine transcription start site of *NRIP-1*. The locations of all binding sites in relation to genes can be found in Table S1.

The ER binding sites adjacent to *TFF-1*, *XBP-1*, *SOD-1*, *NRIP-1*, and *DSCAM-1* were validated by ER ChIP and standard PCR (Figures 2A–2E). Also, quantitative PCR was performed on each of these sites after ER ChIP (Figure 2F), confirming these putative in vivo binding sites as genuine ER binding sites. To test whether these discrete ER recruitment regions were unique to estrogen action in MCF-7 cells, we performed ER ChIP and directed real-time PCR against the same sites in T47-D breast cancer cells. These data confirmed that the majority of the sites identified in MCF-7 cells were also regions of estrogen-dependent ER binding in a second ER-positive breast cancer cell line (data not shown), highlighting the conservation of specific ER-chromatin association sites.

A Significant Number of ER Binding Sites Reside Adjacent to Estrogen Gene Targets

Estrogen-mediated transcript changes were identified by converting RNA from vehicle or estrogen-stimulated MCF-7 cells into double-stranded cDNA and hybridizing to the chromosome 21 and 22 tiled microarrays. Thirty-five genes (4.4% of all genes) appeared to be transcribed, after which real-time primers were made against all these transcripts and quantitative RT-PCR showed that 12 transcripts on chromosomes 21 and 22 were estrogen induced (Table 1). Eleven of these twelve genes had ER binding clusters within 200 kb. The only estrogen-regulated gene that did not have an adjacent ER binding cluster was *ATP5J*. *TFF-1*, *XBP-1*, and *NRIP-1* were in the small list of 1.5% of genes upregulated following estrogen stimulation (Supplemental Raw Data). *DSCAM-1* and *SOD-1* were not upregulated by estrogen stimulation at the 3 hr time point assessed but were transcribed after 6 hr of estrogen stimulation, as determined by RT-PCR (Figure S2). This delay between ER association and transcription of *DSCAM-1* and

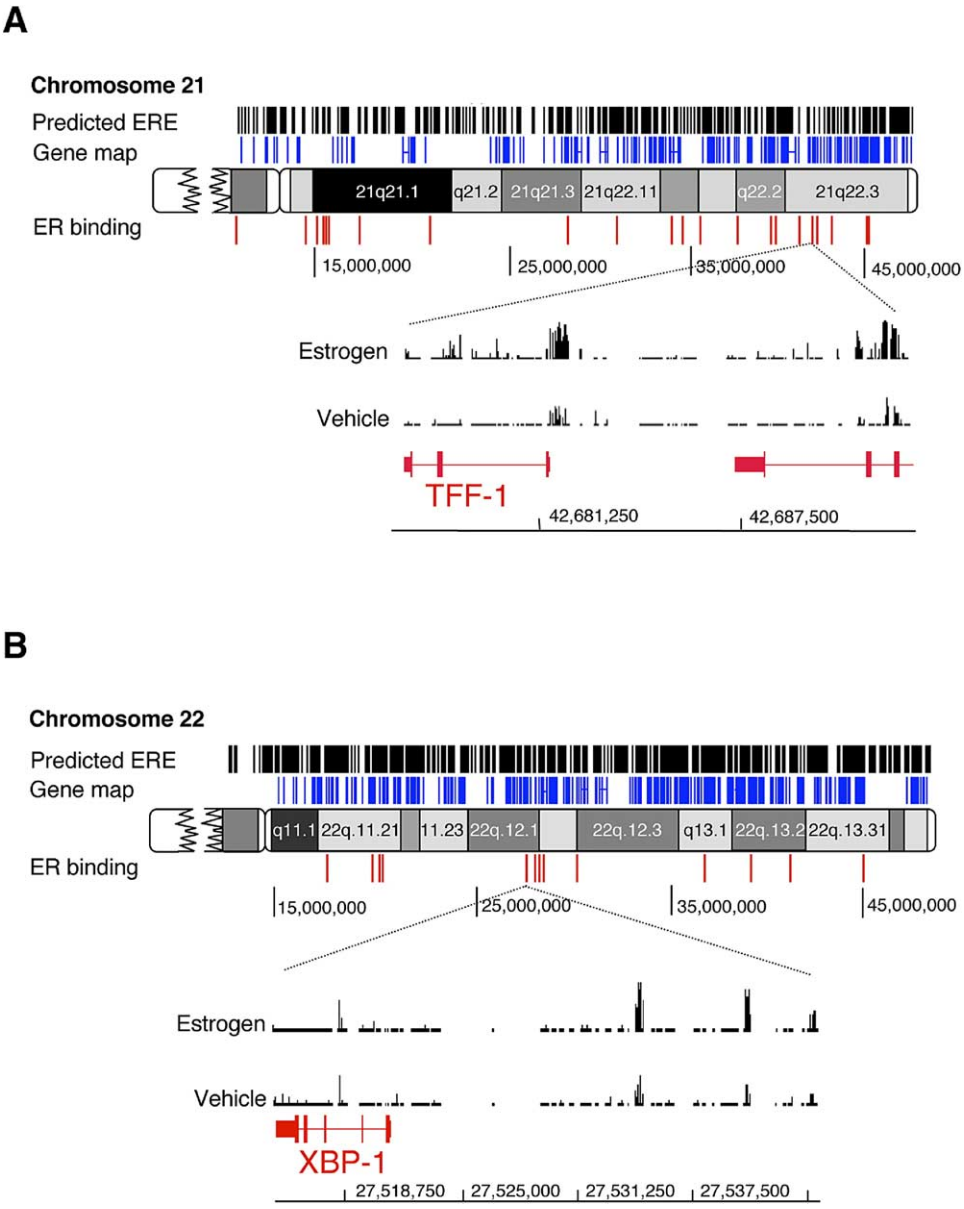

Figure 1. Map of ER Binding Sites on Chromosomes 21 and 22 after Estrogen Stimulation

The visual representation of ER binding sites on chromosomes 21 (A) and 22 (B) are shown. Gene locations are shown in blue bars. Gene locations are based on the April 2003 genome freeze in the UCSC browser using Genbank RefSeq positions. Predicted EREs are shown as black bars and ER binding sites are shown as red bars.

(A) An expanded view of the *TFF-1* gene region is shown as signal difference between ER ChIP and Input DNA for both the estrogen- and vehicle-treated cells. The *TFF-1* gene is shown in its genuine 3′-5′ orientation. The gene adjacent to *TFF-1* is not an estrogen target.

(B) Expanded view of the *XBP-1* gene region on chromosome 22. The *XBP-1* gene is shown in its genuine 3′-5′ orientation.

SOD-1 may be a consequence of a requirement for subsequent modification of the receptor complex or the requirement for the production of other factors involved in ER action but not necessarily part of an ER complex. Regardless of the mechanism for the transcriptional delay, it now appears that early and at least some delayed estrogen-regulated genes recruit the receptor with the same kinetics. This implies that events subsequent to ER binding are responsible for timing the initiation of transcription of these delayed targets.

Distal ER Binding Domains Function as Transcriptional Enhancers

The significant sequence distance between many of the ER binding sites and the putative target gene complicates their functional validation. However, we explored the possibility that these ER binding sites may recruit components indicative of transcriptional activation. RNA PolII ChIP followed by real-time PCR was performed on a subset of the putative regulatory regions adjacent to *TFF-1*, *XBP-1*, *DSCAM-1*, *NRIP-1*, and

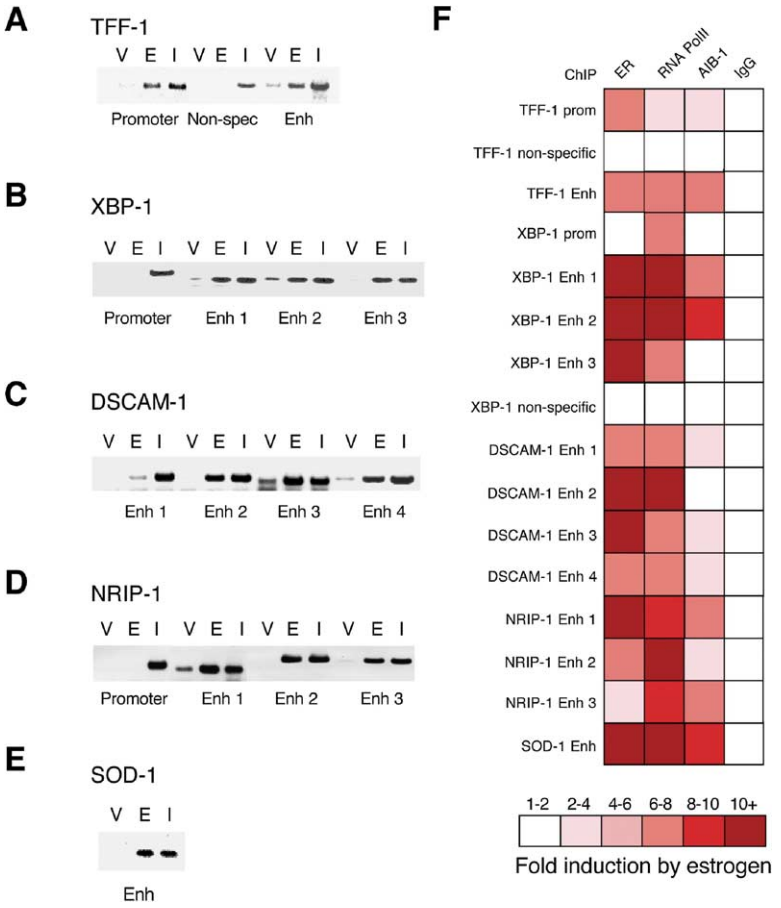

Figure 2. Validation of the In Vivo Binding of the Transcription Complex to Regulatory Regions

ChIP of ER and standard PCR of sites adjacent to *TFF-1* (A), *XBP-1* (B), *DSCAM-1* (C), *NRIP-1* (D), and *SOD-1* (E). *TFF-1* nonspecific and *XBP-1* promoter primers were included as negative controls. The lanes are vehicle (V), estrogen (E), and Input (I).

(F) ChIP of ER, RNA PolII, AIB-1, or IgG control and real-time PCR of binding regions. The data are estrogen-mediated fold enrichment compared to vehicle (ethanol) control. The color intensity reflects the fold change as described in the legend. *TFF-1* nonspecific and *XBP-1* nonspecific primers were included as negative controls. The data are the average of three replicates ± SD.

SOD-1 genes. Interestingly, RNA PolII association was seen with all of these sites in an estrogen-dependent manner (Figure 2F). Furthermore, ChIP of AIB-1, an oncogenic ER coactivator (Kuang et al., 2004; Torres-Arzayus et al., 2004), confirmed that AIB-1 is also present on all of these "regulatory" sites following estrogen exposure (Figure 2F). As negative controls, primers were designed against the intergenic region between the *TFF-1* promoter and enhancer and against a region 7 kb from *XBP-1* enhancer 3. Neither ER nor any of the other factors were found associated with these control regions. In addition, we examined the promoter of *XBP-1*. Although ER protein association was not observed at the *XBP-1* promoter, RNA PolII was found enriched at this site supporting the hypothesis that *XBP-1* is transcriptionally activated by ER.

To explore the possibility that the distal enhancer regions not only function as sites of protein recruitment but physically play a role during transcription of the adjacent gene, we performed a chromosome capture assay (Dekker et al., 2002) to assess whether promoter and enhancer sequences were components of the same chromatin regions. Hormone-depleted MCF-7 cells were stimulated with vehicle or estrogen, and the fixed chromatin was digested with a specific restriction enzyme (BtgI), followed by ER ChIP and ligation. After ligation, the ligated chromatin mix was washed and the crosslinking was reversed. One primer in the *TFF-1* promoter and one primer in the *TFF-1* enhancer were used to PCR potentially ligated fragments of DNA (Horike et al., 2005). As seen in Figure 3A, *TFF-1* promoter and enhancer DNA was ligated together only in the presence of estrogen, confirming that estrogen-mediated transcription of *TFF-1* involves direct physical interaction between the enhancer and promoter. No interaction was seen in the no-digestion control or no-ligation control. We performed the same experiment using the BsmI restriction enzyme that cuts the genuine *NRIP-1* promoter (as determined by 5′ RACE) and enhancer 3 region. Remarkably, after ligation, we were able to PCR a 1 kb fragment that corresponded to the ligated promoter-enhancer regions using one promoter-specific and one enhancer-specific primer (Figure 3B). This estrogen-dependent interaction of the distal (144 kb) ER binding site with the promoter of the *NRIP-1* gene confirms the authenticity of these distal sites as transcriptional regulatory domains.

The finding that RNA PolII is recruited to the majority of ER binding sites, even those removed from known transcription sites, led us to investigate the possibility that these binding sites can function as genuine enhancers. To this end, we cloned 23 ER sites (40% of all ER binding sites) into a pGL-3 luciferase vector containing an SV40 promoter and transfected these vectors into hormone-depleted MCF-7 cells which where subsequently treated with estrogen or vehicle control. pGL-3 empty vector was used as a negative control, and transfections were normalized with pRL null. Al-

Table 1. List of ER Binding Site Clusters and Relative Locations to Putative Gene Targets

Cluster Number	Number of Binding Sites	Start	Stop	Closest Regulated Gene
1	1	21: 10048850	10049271	
2	1	14600251	14600737	
3	1	15171656	15172273	
4	6	15467150	15738864	NRIP-1
5	1	17422343	17422868	
6	1	21532885	21533421	
7	1	29151881	29152882	
8	1	31821967	31822715	SOD-1
9	2	35021165	35027898	
10	1	35510057	35510719	
11	2	36480740	36487032	
12	1	38635468	38636783	
13	10	40363341	40675801	DSCAM-1
14	1	41911683	41912284	
15	1	42005946	42006169	PRDM15
16	2	42680784	42691725	TFF-1
17	1	42830736	42831350	
18	1	43564518	43565261	NDUFV3
19	2	45606461	45663897	
20	1	45790004	45790654	Col18A1
21	2	22: 17159455	17194014	
22	1	19566341	19566809	
23	3	19822950	19945255	
24	3	27534171	27543908	XBP-1
25	1	28106122	28107112	AP1B1
26	1	28237489	28238464	
27	1	28519139	28520023	
28	2	30300284	30307434	PISD
29	2	37030766	37033295	
30	1	39371665	39372232	
31	1	41361325	41361720	Predicted
32	1	45100090	45100552	

The 32 transcriptional clusters are shown, with the start and stop locations of the ER binding sites.

most 75% of the ER binding domains contained estrogen-induced enhancer characteristics in an in vitro transcription model (Figure 3C), supporting the hypothesis that the distal binding sites play transcriptional regulatory roles.

ER Binding Sites Are Conserved Across Species

To identify if the ER binding sites are conserved between human and mouse genomes, we assessed the identity in sequence in a window of 6 kb from the center of all 57 ER binding sites. This conservation was mapped within a 500 bp window at a single nucleotide resolution and confirms a strong conservation at the center of the ER binding site and the 500 bp on either side of the middle of the peak (Figure 4A). However, conservation decreased to background levels at a distance of 1 kb or more from the center of the ER binding sites. This supports the hypothesis that the discrete ER binding sites we see in MCF-7 cells are conserved between species and likely play a more general role in ER action in other cellular systems.

A Screen for Common Sequences Enriched in Genuine ER Binding Regions Suggests the Importance of Forkhead Factors in Estrogen Action

An unbiased search for common sequence motifs (Liu et al., 2002) within the 57 individual ER binding sites on chromosomes 21 and 22 revealed the significant recurrence of two motifs. A consensus 15 base sequence identical to the canonical ERE was present in 49% of all the ER binding sites on chromosomes 21 and 22 (Figure 4B; Klinge, 2001). The likelihood of an ERE occurring in one of the ER binding sites was significantly increased when compared to all of chromosomes 21 and 22 ($p = 1.33E-15$). In the ER binding sites lacking a canonical ERE, a majority were found to contain one or more ERE half-sites, and the occurrence of ERE half-sites was also nonrandom ($p = 2.16E-14$). To confirm that our failure to find ER binding at other EREs (5500 predicted EREs on chromosomes 21 and 22, as listed in Figures 1A and 1B) was not due to the insensitivity of our ChIP-microarray technique, we performed ChIP for ER followed by PCR for several randomly selected, predicted but nonfunctional perfect EREs on chromosomes 21 and 22. No ER association was found at any of these sites (data not shown).

We next determined whether DNA sequences other than the classical ERE were found at the ER binding sites by analyzing the bound sequences for conserved motifs after removing the EREs. This analysis revealed the presence of a Forkhead factor binding site in 54% of the 57 ER binding regions (Figure 4B), a finding that would only occur by chance with a probability of $p = 1.23E-8$. Forkhead binding motifs were found in 56% of the ER binding regions that contain a canonical ERE. Using the consensus Forkhead motif recurring within these regions (Figure 4B), we determined the probability of this motif residing within predicted ERE regions that are not bound by ER in vivo (18.45%). This significant enrichment of a Forkhead motif within ER binding regions ($p = 3.78E-7$) suggested the presence of adjacent Forkhead motifs may play a role in determining ER binding. The finding that the largest category of sites contains both an ERE and a Forkhead motif (47.4%) strongly suggests a functional interaction (Figure 4C).

Forkhead Proteins Play a Combinatorial and Essential Role in ER Binding and ER-Mediated Gene Transcription

A combinatorial interaction between Forkhead and ER pathways has been previously suggested for a small number of specific genes. HNF-3α (FoxA1) Forkhead binding domains within the promoter of the estrogen-regulated genes *TFF-1* (Beck et al., 1999) and Vitellogenin B1 (Robyr et al., 2000) have been shown to be important for gene transcription, and they have been shown to interact directly with ER in yeast two-hybrid experiments (Schuur et al., 2001). The function of Forkhead proteins can be regulated by their nuclear-cytoplasmic distribution depending on their phosphorylation (Brunet et al., 1999; Kops et al., 1999). We therefore determined that FoxA1 localized to the nucleus before and after estrogen stimulation of MCF-7 cells (data not shown).

We next determined whether FoxA1 was recruited along with ER to the ER binding domains. Directed ChIP of FoxA1 followed by real-time PCR of all 57 ER binding regions on chromosomes 21 and 22 revealed a high degree of concordance between regions that recruit ER and FoxA1. Approximately 48% of all of the ER

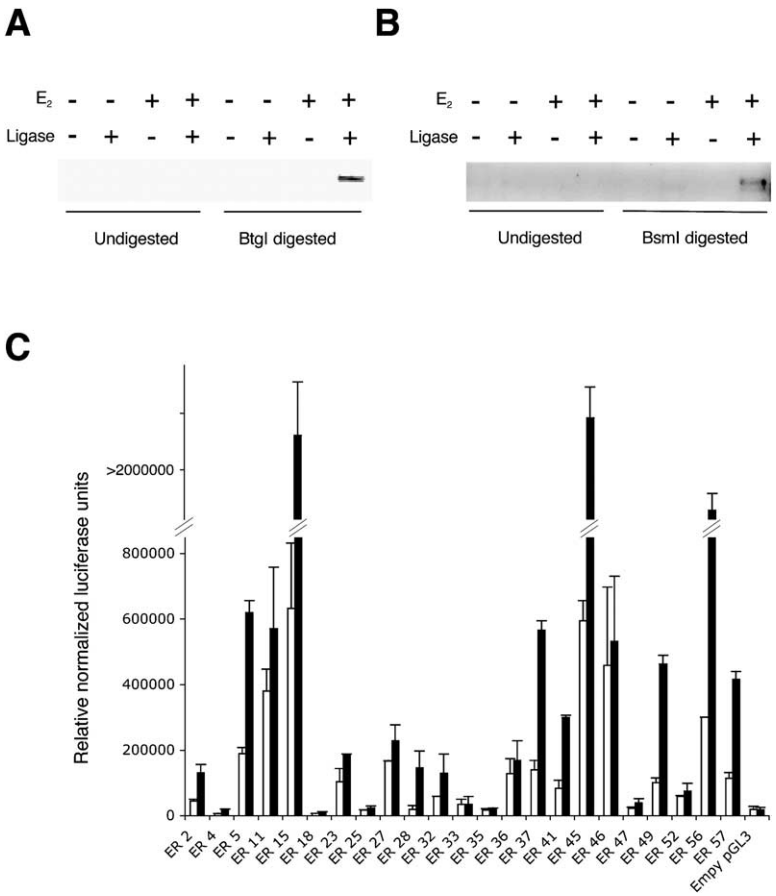

Figure 3. Interaction of Promoter-Enhancer Domains and Transcriptional Activity of Enhancer Regions

(A) Chromosome capture assay was performed after digesting fixed chromatin from vehicle- or estrogen-treated cells with the BtgI restriction enzyme. Primers flanking the *TFF-1* promoter and enhancer were used to amplify DNA after ligation. Undigested controls and no ligase controls were included.

(B) Chromatin was digested with BsmI, and one primer flanking the *NRIP-1* promoter and one in enhancer 3 region were used to amplify a specific product after ligation.

(C) ER binding sites were cloned into the pGL-3 promoter vector and transfected into hormone-depleted MCF-7 cells, after which vehicle (open bars) or estrogen (solid bars) was added. Empty pGL3-promoter vector was used as a negative control. Cotransfection of pRL null Renilla vector was included as a normalizing control. The data are the average of three replicates ± SD.

binding domains showed FoxA1 interaction, although the pattern of recruitment differed from site to site (Figure S3). A majority of the regions containing FoxA1 did so in the absence of estrogen, but FoxA1 binding was decreased following estrogen stimulation. This was the case for *NRIP-1* enhancer 1, *DSCAM-1* enhancer 1, and *TFF-1* promoter (Figure 5A). FoxA1 association with *XBP-1* enhancer 2 was clearly observed but was not diminished after estrogen addition (Figure 5A). All of these ER binding sites contained a Forkhead motif and an ERE or ERE half-site (Figure 5B). FoxA1 was not seen to bind to *XBP-1* enhancer 3, which lacks a Forkhead motif (Figure 5). However, several regions containing Forkhead motifs did not recruit FoxA1, and several ER binding domains that lacked Forkhead motifs did bind FoxA1. This complex interplay between FoxA1, ER, and binding sites within chromatin likely involves adjacent regions to the ER binding sites and may involve other proteins. Despite this, it is clear that a significant proportion of ER binding sites, especially those adjacent to actively transcribed genes, contain FoxA1 prior to estrogen stimulation and ER recruitment to the same regions.

To determine the importance of FoxA1 in mediating ER association with chromatin, we developed siRNA to the 3′UTR of *FoxA1* mRNA. Specific targeted knockdown of FoxA1 protein was achieved (Figure 6A), without changes in control protein or ER protein levels (data not shown). A luciferase siRNA (siLuc) was used as a negative control. MCF-7 cells were deprived of hormones for 24 hr and siLuc, or siRNA to *FoxA1*, was transfected for 6 hr, after which hormone-depleted media was added for a further 48 hr and cells were stimulated with estrogen or vehicle. ER ChIP and real-time PCR of a number of previously validated binding sites was performed. The decrease in *FoxA1* completely impeded the ability of ER to bind to *TFF-1* promoter, *XBP-1* enhancer 1, and *NRIP-1* enhancer 2 (Figure 6B), as well as *DSCAM-1* enhancer 1 (data not shown). No changes were observed on the *XBP-1* promoter, which functioned as a negative control (Figure 6B).

Since the targeted knockdown of FoxA1 inhibited the ability of ER to associate with in vivo ER binding sites, we assessed the effect of Forkhead downregulation on estrogen-mediated transcription. After siLuc or siFoxA1 transfection, cells were stimulated with estrogen or vehicle for 6 hr and mRNA changes in all 12 estrogen target genes on chromosomes 21 and 22 were assessed. The estrogen-induced increases in all 12 estrogen targets were abolished when FoxA1 was downregulated (Figure 6C), but no changes were observed in *GAPDH* control mRNA levels. The essential role for the FoxA1 Forkhead protein during transcription of all estrogen target genes on chromosomes 21 and 22 confirms a general requirement of FoxA1 for ER transcription.

Discussion

A complete picture of ER-mediated gene activation has begun to emerge in recent years, with a coordinated

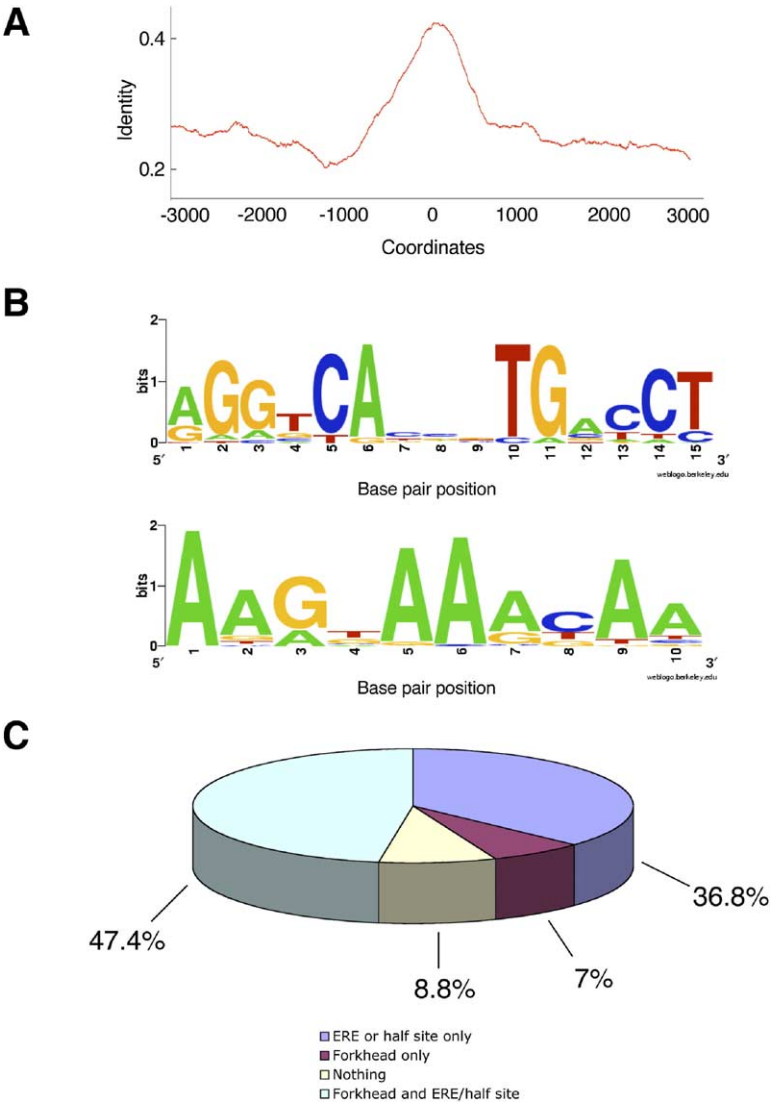

Figure 4. Conservation of ER Binding Sites and Presence of Enriched Motifs

(A) Sequence homology of ER binding sites and surrounding sequence between human and mouse genomes. The center of ER peaks is designated coordinate 0.

(B) An unbiased motif screen of all the ER binding sites on chromosomes 21 and 22 revealed the presence of two enriched motifs, an ERE and a Forkhead binding motif, both of which are visually represented in WebLogo (http://weblogo.berkeley.edu).

(C) The occurrence of ERE or ERE half-sites and Forkhead sites within the 57 ER binding sites on chromosomes 21 and 22.

and timely cycling of receptor, nuclear coactivators, chromatin remodelling proteins, and the transcription machinery on and off target promoters (Metivier et al., 2003; Shang et al., 2000). However, these studies oversimplify the problem by focusing on the promoter proximal region of one or two target genes and largely ignore the remaining chromosomal sequence. Here, we have interrogated the association of ER across entire chromosomes, including intergenic regions that contain potential *cis*-regulatory domains. These ChIP-microarray experiments demonstrate the ability to identify genuine in vivo ER protein binding sites in previously unexplored regions of the genome. Interestingly, while a few of the ER binding sites were found directly adjacent to ER target genes, most were found at significant distances including several >100 kb removed from transcription start sites. Of the 57 ER binding sites (within 32 potential transcriptional regulatory clusters), only a very small number of proximal promoters recruited ER, despite the fact that the other genes were estrogen induced. The presence of multiple components of the transcriptional machinery at distal sites combined with the ability of chromosome conformation capture assays to demonstrate that these distant sites are physically associated with promoter-proximal regions suggests that they play an important role in estrogen-mediated regulation.

A significant volume of work has focused on identifying essential domains within the proximal promoters of known estrogen regulated genes (Dubik and Shiu, 1992; Petz et al., 2002; Porter et al., 1996; Teng et al., 1992; Umayahara et al., 1994; Vyhlidal et al., 2000; Weisz and Rosales, 1990). The conclusions drawn from this large volume of data implicate a number of motifs, including Sp1, AP-1, and GC-rich regions as important *cis*-regulatory domains in ER-mediated transcription. However, our data demonstrate ER regulatory sites at distances several orders of magnitude greater than was focused on in the past, suggesting that they may function in ways analogous to the β-*globin* LCR (Sawado et al., 2003).

Nonbiased motif scanning of the genuine in vivo ER binding sites identified a canonical ERE in the majority of ER binding sites that represented only 1.5% of EREs

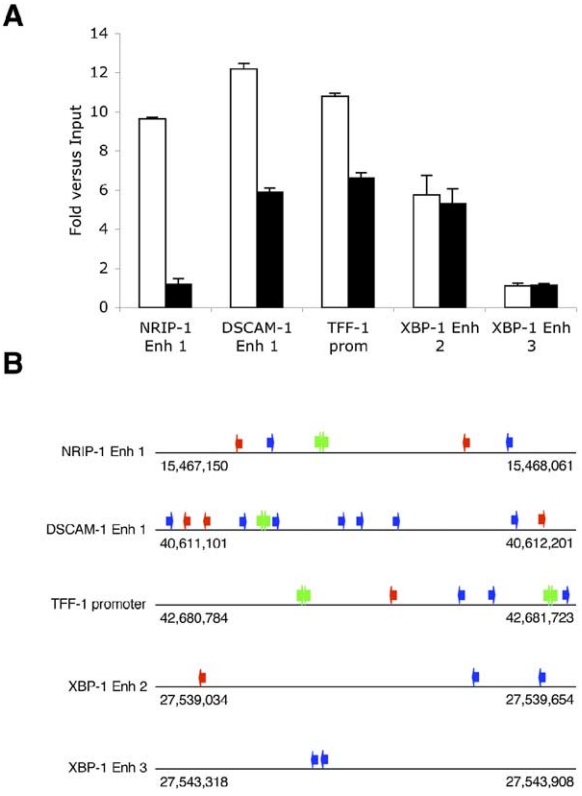

Figure 5. Recruitment of Forkhead Protein to ER Binding Domains

(A) ChIP of FoxA1 followed by real-time PCR of *NRIP-1* enhancer 1, *DSCAM-1* enhancer 1, *TFF-1* promoter, and *XBP-1* enhancer 2. *XBP-1* enhancer 3 is included as a control which does not recruit FoxA1. Data are shown as fold change versus input and are the average of three replicates ± SD. Open bars are vehicle treated and solid bars are estrogen treated.

(B) Schematic diagram showing the relative location of ERE motifs (inverted green arrows), ERE half-sites (blue arrows), and Forkhead motifs (red arrows). Chromosome nucleotide locations are given.

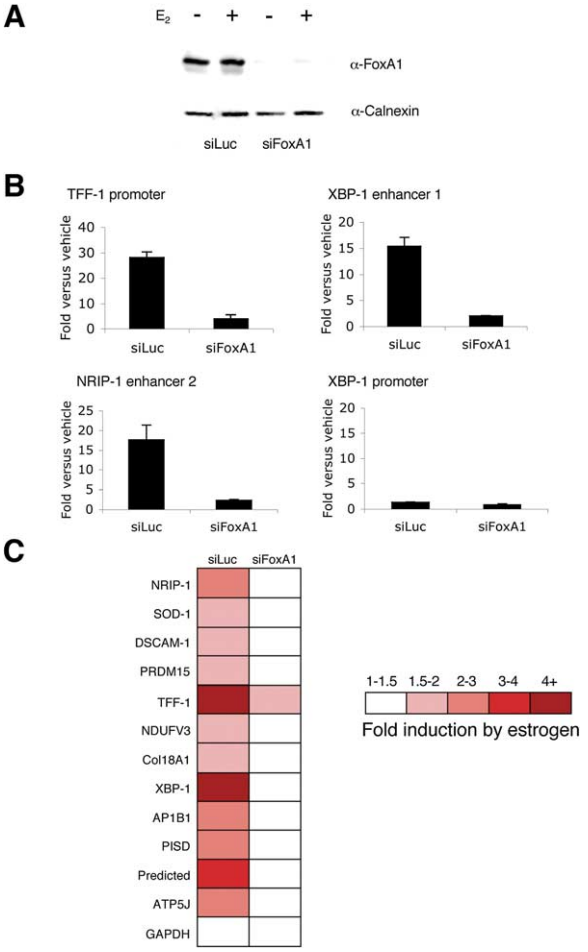

Figure 6. Specific Targeted Knockdown of FoxA1 and the Effects on Estrogen-Mediated Transcription

(A) siRNA to FoxA1 was transfected into hormone-depleted MCF-7 cells, and changes in FoxA1 protein levels were determined after vehicle or estrogen treatment. SiLuc was used as a transfection control and Calnexin was used as a loading control.

(B) ER ChIP was performed after vehicle or estrogen treatment of siLuc or siFoxA1 transfected cells and real-time PCR was conducted on *TFF-1* promoter, *XBP-1* enhancer 1, *NRIP-1* enhancer 2, as well as *XBP-1* promoter as a negative control. The data are fold enrichment over vehicle-treated.

(C) Changes in mRNA levels of all estrogen-regulated genes on chromosomes 21 and 22 after siLuc or siFoxA1. The data are estrogen-mediated fold enrichment compared to vehicle (ethanol) control and are the average of three separate replicates ± SD. The color intensity reflects the fold change as described in the legend.

predicted by bioinformatics alone. Previous approaches for motif identification involved computational-based methods for identifying response elements, after which gene proximal sites are included as potential binding domains (Bajic and Seah, 2003; Bourdeau et al., 2004). The current data suggest that while ER binding involves interaction with consensus ERE motifs, the presence of such motifs is insufficient to dictate receptor-chromatin association. Furthermore, the exclusion of response elements further than several kilobases from transcription start sites eliminates distal regulatory regions that may be the primary receptor-chromatin interaction sites.

Since the presence of an ERE alone is insufficient to define an authentic ER regulatory site, we searched for other conserved sequences and found that Forkhead factor binding sites are present near authentic EREs significantly more frequently than those that do not bind ER. We showed that a Forkhead factor (FoxA1) binding was essential for ER-chromatin interactions and subsequent expression of estrogen gene targets. A link between *ER* and *FoxA1* has previously been shown, with their expression correlated in breast cancer cell lines (Lacroix and Leclercq, 2004). FoxA1 protein can bind condensed chromatin via its winged-helix DNA binding domains that mimic histone linker proteins (Cirillo et al., 2002; Cirillo et al., 1998). Unlike histone proteins however, FoxA1 does not contain the amino acid composition to condense chromatin and it therefore is thought to promote euchromatic conditions. As such, it is possible that the presence of FoxA1 identifies specific regions within chromatin to facilitate the association of the ER transcription complex. Our data suggest that FoxA1 is present on the chromatin at a number of regions, after which ER can associate with these spe-

cific sites. Downregulation of FoxA1 inhibits the ability of ER to associate with its binding sites, confirming the requirement for Forkhead-directed association of ER with chromatin, despite the fact that these sites contain sufficient information, in the form of an ERE, for ER docking. This, combined with a recent investigation showing that FoxA1 can directly modulate chromatin in the MMTV promoter and can positively enhance transcription by the glucocorticoid receptor (Holmqvist et al., 2005), supports a general model for FoxA1 involvement in nuclear receptor transcription.

We have taken an unbiased approach to identify regions of chromatin, both promoter proximal and intergenic sequences, which are involved in ER-mediated transcriptional activity. We find a limited number of bona fide ER binding sites on chromosomes 21 and 22, with a significant enrichment of canonical ERE palindromes and half-sites within the binding sites. Moreover, the presence of Forkhead binding motifs and the subsequent identification of a functional role for the Forkhead protein FoxA1 in estrogen signaling exemplifies the power of this approach to identify important regulatory domains within the vast regions of unexplored sequence of the human genome.

Experimental Procedures

Chromatin Immunoprecipitation (ChIP)-Microarray Preparation

ChIP was performed as previously described (Shang et al., 2000), with the following modifications. Two micrograms of antibody was prebound for a minimum of 4 hr to protein A and protein G Dynal magnetic beads (Dynal Biotech, Norway) and washed three times with ice-cold PBS plus 5% BSA and then added to the diluted chromatin and immunoprecipitated overnight. The magnetic bead-chromatin complexes were collected and washed six times in RIPA buffer (50 mM HEPES [pH 7.6], 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5 M LiCl). Elution of the DNA from the beads was as previously described (Shang et al., 2000). Antibodies used were as follows: ERα (Ab-10) from Neomarkers (Lab Vision, United Kingdom), ERα (HC-20), RNA PolII (H-224), AIB-1/RAC3 (C-20), HNF-3α/FoxA1 (H-120), mouse IgG (sc-2025), and rabbit IgG (sc-2027) from Santa Cruz (Santa Cruz Biotechnologies, California). Ligation-Mediated PCR was performed as previously described (Ren et al., 2002). Labeling was performed as previously described (Kapranov et al., 2002). Microarrays used were Affymetrix Genechip chromosome 21/22 tiling set P/N 900545.

Data Analysis

1,054,325 probe pairs were mapped to chromosomes 21 and 22 according to the NCBIv33 GTRANS Libraries provided by Affymetrix. (PM-MM) value was recorded for each probe pair, and a probe pair was removed if either PM or MM was flagged as outlier by the Affymetrix GCOS software. The samples (three ER+ ChIP and three genomic inputs) were normalized by quantile normalization (Bolstad et al., 2003) based on a combined 76 ChIP experiments obtained from public domain and Dana-Farber Cancer Institute. The behavior of every probe pair i, assumed to be $N(\mu_i, \sigma_i^2)$, was estimated from the 76 normalized experiments. A two-state (ChIP-enriched state and nonenriched state) Hidden Markov Model with the following parameters was applied to each sample to estimate the probability of ChIP enrichment at each probe pair location:

Transition probabilities: 300/1,054,325 for transition to a different state,
1 – (300/1,054,325) for staying in the same state.
Emission probabilities: $N(\mu_i, \sigma_i^2)$ for nonenriched hidden state,
$N(\mu_i + 2\sigma_i, (1.5\sigma_i)^2)$ for enriched hidden state.

To combine the results from the six samples, an enrichment score was calculated as the average enrichment probability in the three ER+ ChIP samples subtracted by the average enrichment probability in the three genomic input samples. Since the tiling array has one 25-mer probe in every 35 bp of nonrepeat regions, the coverage of every probe was extended by 10 bp on both ends. An enriched regions is defined as run of probes with enrichment score >50% and covering at least 125 bp. Each enriched region can tolerate up to two neighboring probes with enrichment score between [10%, 50%]. If two neighboring probes are more than 210 bp apart, the enriched region is broken into two separate blocks. A summary enrichment score was obtain for each enriched region, which is the enrichment score summation for all the probes in the region divided by the square root of the number of probes in the region. This summary enrichment score represents the relative confidence of a predicted enriched region.

Sequence Analysis

The genomic DNA of every ChIP-enriched region was retrieved from UCSC genome browser and ranked by the summary enrichment score. MDscan algorithm (Liu et al., 2002) was applied to the sequences to find enriched sequence pattern that is the putative estrogen receptor binding motif. To find a motif of width w, MDscan first enumerates each w-mer in the highest ranking sequences and collects other w-mers similar to it in these sequences to construct a candidate motif as a probability matrix. A semi-Bayes scoring function was used to remove low-scoring candidate motifs and refine the rest by checking all w-mers in all the ChIP-enriched sequences. A high-scoring motif (with similar consensus) consistently reported multiple times at different motif widths indicates a strong prediction.

We expanded all 57 of the ER binding sites equally in each direction to have a length of 6 kb. The human-mouse conservation score of each nucleotide in the expanded binding region is defined as the average sequence identity (# matched nucleotides – # indels)/500 of a 500-mer window centered at the nucleotide. The human (hg15) /mouse (mm3) BLASTZ (Schwartz et al., 2003) genome alignments were downloaded from http://genome.ucsc.edu.

Real-Time PCR

Primers were selected using Primer Express (Applied Biosystems). Five microliters of precipitated and purified DNA was subjected to PCR using the Applied Biosystems SYBR Green Mastermix. Relative DNA quantities were measured using the PicoGreen system (Molecular Probes, Oregon). All primer sequences and locations are listed in Table S2.

Double-Stranded cDNA Synthesis

Total RNA was converted to double stranded cDNA according to the Invitrogen Superscript double-stranded cDNA synthesis manufacturer's instructions. The RNA was primed with 250 ng oligo(dT) (Invitrogen) and 25 ng random hexamers (Gibco). cDNA was fragmented and labeled as described above.

5′RACE

5′ RACE was performed according to the manufacturer's instructions (Invitrogen). The primers sequences used were as follows: *NRIP-1* RT primer (5′-TGCCTGATGCATTAGTAATCC-3′), *NRIP-1* nested primer 1 (5′-GAGCCAAGCTCTTCTCCATGAGTCATGTTC-3′), and *NRIP-1* nested primer 2 (5′-ACCTTCCATCGCAATCAGAGA GAGACGTACTG-3′). The PCR product was cloned and sequenced by standard methods.

Chromosome Capture Assay

Fixed chromatin was digested overnight with specific restriction enzymes after which ER ChIP was set up as described above. After overnight ChIP, the beads were precipitated and resuspended in ligation buffer (NEB, Massachusetts) and overnight ligation was performed. The beads were collected, washed, and the formaldehyde crosslinking was reversed as described above. Primers used to amplify annealed fragments were as described in Table S2.

Luciferase Enhancer Activity

ER binding sites were amplified by PCR and cloned into the pGL-3-promoter vector (Promega). Hormone-depleted MCF-7 cells were transfected with each of the ER binding domain vectors with Lipofectamine 2000 (Invitrogen), and total protein lysate was harvested after estrogen or ethanol addition for 24 hr. Transfections were normalized by the cotransfection of the pRL null renilla luciferase vector and renilla and firefly luciferase activity was assessed using the dual luciferase kit (Promega).

Western Blotting

SDS-PAGE was performed as previously described (Carroll et al., 2000). Antibodies used were FoxA1/HNF-3α (ab5089), from AbCam (Cambridge, United Kingdom) and Calnexin (H-70) from Santa Cruz (California).

Short Interfering (si) RNA

A 21 bp siRNA was designed against the FoxA1 transcript and synthesized by Dharmacon (Lafayette, Colorado). siRNA was transfected using Lipofectamine 2000 (Invitrogen). The siRNA sequences used were as follows: *siFoxA1* sense 5′-GAGAGAAAAAA UCAACAGC-3′ and antisense 5′-GCUGUUGAUUUUUUCUCUC-3′; siLuc sense 5′-CACUUACGCUGAGUACUUCGA-3′ and antisense 5′-UCGAAGUACUCAGCGUAAGUG-3′.

Supplemental Data

Supplemental Data include four figures, two tables, and raw data files and can be found with this article online at http://www.cell.com/cgi/content/full/122/1/33/DC1/.

Acknowledgments

The authors would like to thank Danielle Lynch and Mitch Lazar for helpful discussions and review of the manuscript. We thank Molin Wang, David Harrington, Cheng Li, and Jun S. Liu for their insights and advice on data analysis. This work was supported by grants from the Claudia Adams Barr Program in Innovative Basic Cancer Research (M.B., P.A.S., X.S.L.), the DF/HCC Breast Cancer SPORE Grant, by DOD Breast Cancer Research Program Awards DAMD17-01-1-0222 (M.B.), W81XWH-04-1-0512 (J.S.C.), DAMD17-02-0364 (P.A.S.), by N.I.H. grants K22 HG02488-01A1 (A.S.B.), F32 CA108380-02 (T.R.G.), T90 DK070078-01 (C.A.M.), and by fellowships from the Fondation Recherche Medicale (J.E.), and the HHMI (A.J.S.).

Received: January 27, 2005
Revised: April 6, 2005
Accepted: May 6, 2005
Published: July 14, 2005

References

Augereau, P., Miralles, F., Cavailles, V., Gaudelet, C., Parker, M., and Rochefort, H. (1994). Characterization of the proximal estrogen-responsive element of human cathepsin D gene. Mol. Endocrinol. *8*, 693–703.

Bajic, V.B., and Seah, S.H. (2003). Dragon gene start finder identifies approximate locations of the 5′ ends of genes. Nucleic Acids Res. *31*, 3560–3563.

Beck, S., Sommer, P., dos Santos Silva, E., Blin, N., and Gott, P. (1999). Hepatocyte nuclear factor 3 (winged helix domain) activates trefoil factor gene TFF1 through a binding motif adjacent to the TATAA box. DNA Cell Biol. *18*, 157–164.

Beckman, J.S., Carson, M., Smith, C.D., and Koppenol, W.H. (1993). ALS, SOD and peroxynitrite. Nature *364*, 584.

Berry, M., Nunez, A.M., and Chambon, P. (1989). Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic sequence. Proc. Natl. Acad. Sci. USA *86*, 1218–1222.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Bourdeau, V., Deschenes, J., Metivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J.H., and Mader, S. (2004). Genome-wide identification of high-affinity estrogen response elements in human and mouse. Mol. Endocrinol. *18*, 1411–1427.

Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J., and Greenberg, M.E. (1999). Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. Cell *96*, 857–868.

Bulger, M., Sawado, T., Schubeler, D., and Groudine, M. (2002). ChIPs of the beta-globin locus: unraveling gene regulation within an active domain. Curr. Opin. Genet. Dev. *12*, 170–177.

Carroll, J.S., Prall, O.W., Musgrove, E.A., and Sutherland, R.L. (2000). A pure estrogen antagonist inhibits cyclin E-Cdk2 activity in MCF-7 breast cancer cells and induces accumulation of p130–E2F4 complexes characteristic of quiescence. J. Biol. Chem. *275*, 38221–38229.

Cavailles, V., Dauvois, S., L'Horset, F., Lopez, G., Hoare, S., Kushner, P.J., and Parker, M.G. (1995). Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. EMBO J. *14*, 3741–3751.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499–509.

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., and Zaret, K.S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. *17*, 244–254.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Dubik, D., and Shiu, R.P. (1992). Mechanism of estrogen activation of c-myc oncogene expression. Oncogene *7*, 1587–1594.

Giamarchi, C., Solanas, M., Chailleux, C., Augereau, P., Vignon, F., Rochefort, H., and Richard-Foy, H. (1999). Chromatin structure of the regulatory regions of pS2 and cathepsin D genes in hormone-dependent and -independent breast cancer cell lines. Oncogene *18*, 533–541.

Halachmi, S., Marden, E., Martin, G., MacKay, H., Abbondanza, C., and Brown, M. (1994). Estrogen receptor-associated proteins: possible mediators of hormone-induced transcription. Science *264*, 1455–1458.

Holmqvist, P.-H., Belikov, S., Zaret, K.S., and Wrange, Ö. (2005). FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. Exp. Cell Res. *304*, 593–603.

Horike, S., Cai, S., Miyano, M., Cheng, J.F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat. Genet. *37*, 31–40.

Ikeda, K., Sato, M., Tsutsumi, O., Tsuchiya, F., Tsuneizumi, M., Emi, M., Imoto, I., Inazawa, J., Muramatsu, M., and Inoue, S. (2000). Promoter analysis and chromosomal mapping of human EBAG9 gene. Biochem. Biophys. Res. Commun. *273*, 654–660.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science *296*, 916–919.

Katsanis, N., Ives, J.H., Groet, J., Nizetic, D., and Fisher, E.M. (1998). Localisation of receptor interacting protein 140 (RIP140) within 100 kb of D21S13 on 21q11, a gene-poor region of the human genome. Hum. Genet. *102*, 221–223.

Klinge, C.M. (2001). Estrogen receptor interaction with estrogen response elements. Nucleic Acids Res. *29*, 2905–2919.

Kops, G.J., de Ruiter, N.D., De Vries-Smits, A.M., Powell, D.R., Bos,

J.L., and Burgering, B.M. (1999). Direct control of the Forkhead transcription factor AFX by protein kinase B. Nature *398*, 630–634.

Kuang, S.Q., Liao, L., Zhang, H., Lee, A.V., O'Malley, B.W., and Xu, J. (2004). AIB1/SRC-3 deficiency affects insulin-like growth factor I signaling pathway and suppresses v-Ha-ras-induced breast cancer initiation and progression in mice. Cancer Res. *64*, 1875–1885.

Lacroix, M., and Leclercq, G. (2004). About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. Mol. Cell. Endocrinol. *219*, 1–7.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat. Biotechnol. *20*, 835–839.

Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751–763.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378–1381.

Pedram, A., Razandi, M., Aitkenhead, M., Hughes, C.C., and Levin, E.R. (2002). Integration of the non-genomic and genomic actions of estrogen. Membrane-initiated signaling by steroid to transcription and cell biology. J. Biol. Chem. *277*, 50768–50775.

Petz, L.N., Ziegler, Y.S., Loven, M.A., and Nardulli, A.M. (2002). Estrogen receptor alpha and activating protein-1 mediate estrogen responsiveness of the progesterone receptor gene in MCF-7 breast cancer cells. Endocrinology *143*, 4583–4591.

Porter, W., Wang, F., Wang, W., Duan, R., and Safe, S. (1996). Role of estrogen receptor/Sp1 complexes in estrogen-induced heat shock protein 27 gene expression. Mol. Endocrinol. *10*, 1371–1378.

Rabiner, L.R. (1989). A tutorial on Hidden Markov models and selected applications in speech recognition. Proc. IEEE *77*, 257–286.

Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. (2002). E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev. *16*, 245–256.

Robyr, D., Gegonne, A., Wolffe, A.P., and Wahli, W. (2000). Determinants of vitellogenin B1 promoter architecture. HNF3 and estrogen responsive transcription within chromatin. J. Biol. Chem. *275*, 28291–28300.

Sawado, T., Halow, J., Bender, M.A., and Groudine, M. (2003). The beta-globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. Genes Dev. *17*, 1009–1018.

Schiff, R., Reddy, P., Ahotupa, M., Coronado-Heinsohn, E., Grim, M., Hilsenbeck, S.G., Lawrence, R., Deneke, S., Herrera, R., Chamness, G.C., et al. (2000). Oxidative stress and AP-1 activity in tamoxifen-resistant breast tumors in vivo. J. Natl. Cancer Inst. *92*, 1926–1934.

Schuur, E.R., Loktev, A.V., Sharma, M., Sun, Z., Roth, R.A., and Weigel, R.J. (2001). Ligand-dependent interaction of estrogen receptor-alpha with members of the forkhead transcription factor family. J. Biol. Chem. *276*, 33554–33560.

Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res. *13*, 103–107.

Shang, Y., and Brown, M. (2002). Molecular determinants for the tissue specificity of SERMs. Science *295*, 2465–2468.

Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A., and Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. Cell *103*, 843–852.

Singh, R.J., Karoui, H., Gunther, M.R., Beckman, J.S., Mason, R.P., and Kalyanaraman, B. (1998). Reexamination of the mechanism of hydroxyl radical adducts formed from the reaction between familial amyotrophic lateral sclerosis-associated Cu,Zn superoxide dismutase mutants and H2O2. Proc. Natl. Acad. Sci. USA *95*, 6675–6680.

Teng, C.T., Liu, Y., Yang, N., Walmer, D., and Panella, T. (1992). Differential molecular mechanism of the estrogen action that regulates lactoferrin gene in human and mouse. Mol. Endocrinol. *6*, 1969–1981.

Torres-Arzayus, M.I., De Mora, J.F., Yuan, J., Vazquez, F., Bronson, R., Rue, M., Sellers, W.R., and Brown, M. (2004). High tumor incidence and activation of the PI3K/AKT pathway in transgenic mice define AIB1 as an oncogene. Cancer Cell *6*, 263–274.

Umayahara, Y., Kawamori, R., Watada, H., Imano, E., Iwama, N., Morishima, T., Yamasaki, Y., Kajimoto, Y., and Kamada, T. (1994). Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. J. Biol. Chem. *269*, 16433–16442.

Vyhlidal, C., Samudio, I., Kladde, M.P., and Safe, S. (2000). Transcriptional activation of transforming growth factor alpha by estradiol: requirement for both a GC-rich site and an estrogen response element half-site. J. Mol. Endocrinol. *24*, 329–338.

Wang, D.Y., Fulthorpe, R., Liss, S.N., and Edwards, E.A. (2004). Identification of estrogen-responsive genes by complementary deoxyribonucleic acid microarray and characterization of a novel early estrogen-induced gene: EEIG1. Mol. Endocrinol. *18*, 402–411.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev. *16*, 235–244.

Weisz, A., and Rosales, R. (1990). Identification of an estrogen response element upstream of the human c-fos gene that binds the estrogen receptor and the AP-1 transcription factor. Nucleic Acids Res. *18*, 5097–5106.